

LEIS DE PARETO PARA GUERRAS E ACTOS DE TERRORISMO

André Marino Barata da Conceição Silva



Mestrado em Engenharia Eletrotécnica e de Computadores
Área de Especialização de Sistemas e Planeamento Industrial
Departamento de Engenharia Electrotécnica
Instituto Superior de Engenharia do Porto

2014

Este relatório satisfaz, parcialmente, os requisitos que constam da Ficha de Disciplina de
Tese/Dissertação, do 2º ano, do Mestrado em Engenharia Electrotécnica e de
Computadores

Candidato: André Marino Barata da Conceição Silva, Nº 1120980, 1120980@isep.ipp.pt

Orientação científica: Carla Pinto, cap@isep.ipp.pt

Co-orientação científica: J. A. Tenreiro Machado, jtm@isep.ipp.pt



Mestrado em Engenharia Eletrotécnica e de Computadores
Área de Especialização de Sistemas e Planeamento Industrial
Departamento de Engenharia Electrotécnica

Instituto Superior de Engenharia do Porto

5 de julho de 2014

Agradecimentos

Gostaria de deixar aqui o meu agradecimento aos meus pais, apoiantes incansáveis durante todo o meu percurso académico, dando-me a motivação necessária para enfrentar os desafios que foram surgindo e encará-los não como uma dificuldade, mas sim como uma oportunidade de me poder melhorar como ser humano.

Também gostaria de agradecer à professora Carla Pinto em toda a orientação que foi dada ao longo desta tese de mestrado, sendo que os conhecimentos anteriores da professora nesta área foram uma mais-valia que tornou esta tese mais rica em conteúdo.

Resumo

Nesta tese estudam-se Leis de Potência (LP), também designadamente Leis de Pareto ou Leis de Zipf. As LP são distribuições estatísticas amplamente usadas na compreensão de sistemas naturais e artificiais. Vilfredo Pareto, no século XIX, foi pioneiro no estudo das LP. Pareto afirma, no seu manual de economia política, “*Cours d’Economie Politique*”, que grande parte da economia mundial segue uma determinada distribuição estatística, em que 20% da população reúne 80% da riqueza total do país, estando, assim uma pequena fração da sociedade a controlar a maior fatia do dinheiro. Isto resume o comportamento de uma variável que segue uma distribuição de Pareto (ou Lei de Potência). Ao longo das décadas que se seguiram, mais investigadores aplicaram as LP a outros fenómenos, como sendo a variação dos rendimentos pessoais ou de empresas, a ocorrência de palavras em textos, as repetições de sons ou conjuntos de sons em composições musicais, o número de vítimas em guerras ou outros cataclismos, a magnitude de tremores de terra, o número de vendas de livros ou CD’s na internet, entre muitos outros.

Este trabalho incide sobre a aplicação das LP a guerras e actos de terrorismo. Estuda-se um conjunto de dados relativos a guerras e actos de terrorismo na Europa e no resto do mundo, ao qual se aplica uma distribuição estatística do tipo LP. É estudado o número de vítimas por região, por período histórico e por tipologia de guerra. Conclui-se que, em todos os casos, os dados são bem aproximados por uma distribuição LP. Há assimetrias a nível regional, estando a Ásia associada a maior severidade, relativamente ao número de vítimas enquanto na Europa e a Austrália as guerras são menos severas. Faz-se ainda uma análise da variação temporal dos parâmetros das diferentes distribuições LP aplicadas. Como trabalho futuro, pretende-se fundamentar teoricamente os resultados e aplicar diferentes distribuições aos dados, como a exponencial e a lognormal, com o objectivo de rejeitá-las comparativamente à LP. Pretende-se ainda explicar como é que factores como a religião, modo de vida, estabilidade política podem influenciar os resultados e alargar o período histórico estudado.

Palavras-Chave

Leis de Potência, Leis de Pareto, Leis de Zipf, guerras, guerrilhas, actos de terrorismo

Abstract

In this study we consider Power Laws (PL), also known as Pareto Laws or Zipf Laws. PL are statistical distributions. Vilfredo Pareto, in the XIX century, was a pioneer in the study of PL. He states, in his manual of political economy, "Cours d'Economie Politique", that a large part of the world economy follows a particular distribution, in which 20% of the population gathers 80% of the total wealth, and so, a small fraction of the society controls the largest share of the money. This summarizes the behavior of a variable that follows a Pareto distribution (or, Power Law). Until now, several researchers applied PL to other natural and human-made phenomena, such as the variation in personal incomes or businesses, the occurrence of words in texts, the repetition of sounds or combinations of sounds in musical compositions, the number of casualties in wars or other cataclysms, the magnitude of earthquakes, the number of sales of books or CD's on the internet, amongst others.

This work focuses on the application of PL to wars and terrorists attacks. It is studied a set of data concerning the number of victims of wars and terrorist attacks in Europe and the rest of the world, and then a statistical distribution of type LP is applied to that data. The number of victims is studied by region, by historical period and by type of war. It is concluded that, in all cases, the data are well approximated by a PL distribution. We identify regional asymmetries in what concerns the number of victims, being Asia associated with more severity, while in Europe and Australia wars are less severe. Moreover, the analysis of the temporal variation of the parameters of different PL distributions is also performed in this work. As a future work we intend to do the theoretically study of the results and to apply distinct distributions, such as exponential and lognormal, namely to reject those comparatively to LP. We also would like to explain how religion, way of life, political stability can influence the results obtained and to enlarge the study to a broader historic period.

Keywords:

Power Laws, Pareto Laws, Zipf Law, war, guerrilla, terrorist attacks.

Índice

AGRADECIMENTOS	I
RESUMO	III
ABSTRACT	V
ÍNDICE	VII
ÍNDICE DE FIGURAS	IX
ÍNDICE DE TABELAS	XI
ACRÓNIMOS.....	XIII
1. INTRODUÇÃO	1
1.1. CONTEXTUALIZAÇÃO	3
1.2. OBJECTIVOS.....	4
1.3. CALENDARIZAÇÃO	4
1.4. ORGANIZAÇÃO DO RELATÓRIO	5
2. LEIS DE POTÊNCIA.....	7
2.1. LEI DE POTÊNCIA	7
2.2. MÉDIA E VARIÂNCIA DA LP	9
2.3. ESTIMAÇÃO DE PARÂMETROS DA LEI DE POTÊNCIA.....	11
2.3.1 ESTIMAÇÃO DE PARÂMETROS: NOTAS FINAIS.....	15
2.4. TESTES PARAMÉTRICOS : GOODNESS-OF-FIT.....	15
2.5. APLICAÇÃO ÀS GUERRAS E ACTOS DE TERRORISMO	16
3. APLICAÇÃO A DADOS REAIS DE GUERRAS E ACTOS DE TERRORISMO.....	19
3.1. NÚMERO DE MORTOS EM GUERRAS A NÍVEL MUNDIAL.....	20
3.2. NÚMERO DE MORTOS EM GUERRAS POR REGIÃO	21
3.3. NÚMERO DE MORTOS EM GUERRAS POR TIPOLOGIA DE GUERRA	27
3.4. NÚMERO DE MORTOS EM GUERRAS POR PERÍODO HISTÓRICO	30
3.5. VARIAÇÃO TEMPORAL DOS PARÂMETROS DAS LP	34
REFERÊNCIAS DOCUMENTAIS	39

Índice de Figuras

Figura 1	Gráfico classificação/frequência antes da normalização	20
Figura 2	Gráfico classificação/frequência após normalização.....	21
Figura 3	Gráfico classificação/frequência após normalização: América do Norte.....	22
Figura 4	Gráfico classificação/frequência após normalização: América Central e Sul	23
Figura 5	Gráfico classificação/frequência após normalização: Europa	23
Figura 6	Gráfico classificação/frequência após normalização: Ásia	24
Figura 7	Gráfico classificação/frequência após normalização: Austrália.....	25
Figura 8	Gráfico classificação/frequência após normalização: África	25
Figura 9	Representação dos parâmetros C e α das guerras por região.....	26
Figura 10	Gráfico classificação/frequência após normalização: Guerrilhas.....	28
Figura 11	Gráfico classificação/frequência após normalização: Religiosas.....	29
Figura 12	Representação dos parâmetros C e α por tipologia de guerra.....	30
Figura 13	Gráfico classificação/frequência após normalização: 1972-1992	31
Figura 14	Gráfico classificação/frequência após normalização: 1992-2002	32
Figura 15	Gráfico classificação/frequência antes da normalização: 2002-2012.....	33
Figura 16	Representação dos parâmetros C e α por período de tempo.....	34
Figura 17	Evolução temporal do parâmetro C	35
Figura 18	Evolução temporal do parâmetro α	35

Índice de Tabelas

Tabela 1	Calendarização do Projeto.....	5
----------	--------------------------------	---

Acrónimos

LP – Leis de Potência

MM – Método dos Momentos

MMV – Método da Máxima Verosimilhança

MMQ – Método dos Mínimos Quadrados

fdp – função densidade de probabilidade

1. INTRODUÇÃO

As Leis de Potência, Leis de Pareto ou Leis de Zipf são distribuições estatísticas que explicam a relação entre duas quantidades, onde uma delas varia segundo uma potência da outra. Por exemplo, o número de cidades com um dado tamanho populacional varia segundo uma potência desse tamanho populacional.

Estas leis foram introduzidas, em 1896, por Vilfredo Pareto no seu manual de economia política “Cours d’Economie Politique” [19]. As LP foram aplicadas inicialmente à economia, para explicar as diferenças de rendimento entre os indivíduos de uma população. De acordo com Pareto, o número de indivíduos com um rendimento maior do que um certo valor x era inversamente proporcional a x . Uma quantidade x obedece a uma LP se for retirada de uma distribuição de probabilidade $p(x) \sim Cx^{-\alpha}$ onde α e C são parâmetros. Na realidade $2 < \alpha < 3$ [8].

A Lei de Pareto, também conhecida como a regra 80/20, é considerada uma “Top-Heavy Distribution”, isto é, os valores mais elevados da variável a analisar estão concentrados numa pequena percentagem da população.

As LP foram adaptadas, dada a sua natureza invariante, a todo um conjunto de fenómenos distintos, nomeadamente à internet, número de habitantes de cidades, fogos florestais, terremotos, guerras, entre outros. No trabalho realizado por Pinto *et al* [20] fez-se uma revisão de vários artigos existentes na literatura sobre LP. Estudaram-se ainda, como exemplo, dados sobre fogos florestais, disponibilizados pelo Departamento de Dados de Protecção da Floresta e Fogos do Estado da Califórnia [5], no período de 2001 e 2010. Foi utilizado, como factor explicativo da frequência de fogos, o tamanho total da área

ardida em acres. Os dados foram tratados e foi construído o gráfico em dupla escala logarítmica nos eixos. Foi possível observar que a distribuição da área ardida segue uma LP, onde a maioria dos fogos possui valores de área ardida muito pequenos, inversamente ao que acontece com um pequeno conjunto de fogos caracterizado por áreas ardidas muito grandes.

D. Rafailidis e Y. Manolopoulos [23], investigadores da universidade grega de Aristotle, analisaram dados relativos a colecções musicais, para identificar a frequência de padrões repetitivos em diferentes compositores. Pretendiam explicar as mudanças das frequências acústicas para diferentes tamanhos de padrões. Um outro objetivo deste estudo foi o de comparar os parâmetros das LP entre os diferentes compositores. Os dados foram recolhidos do site da KernScores [15] e da Multimedia Library [17], sendo considerados 570 trabalhos de Bach, 228 de Correlli e 160 de Haydn. Os resultados obtidos permitiram concluir que os dados recolhidos eram bem aproximados por LP. Em particular, concluíram que a maioria dos padrões repetitivos, para um tamanho superior a três, estava concentrada numa percentagem de peças muito pequena, sendo que as restantes peças tinham padrões repetitivos na generalidade inferiores a 3.

Outra área de aplicação das LP é o tamanho das cidades. Auerbach [1] foi o primeiro investigador a observar que o tamanho das cidades era inversamente proporcional à sua frequência. O seu trabalho foi seguido por Zipf [30]. Outros autores testaram a validade desta teoria, tendo concluído que estas LP só se observavam em cidades com um número de habitantes superior a um dado valor. M.E.J. Newman [18] analisou a distribuição cumulativa de várias cidades dos E.U.A., com dados retirados do Departamento de Censos em 2000. Newman chegou à conclusão de que uma recta de declive negativo se adaptava à distribuição dos dados. Pinto et al [21] estudaram a distribuição da população da Turquia, usando dados dos censos dos últimos 40 anos (retirados do site City Population de Thomas Brinkhoff [7]). As representações gráficas com os eixos logaritmizados permitiram concluir que os dados poderiam ser explicados pela Lei de Zipf.

Nesta tese pretende-se aplicar as LP ao tratamento de dados relacionados com guerras e actos de terrorismo, nomeadamente ao número de vítimas que daí resultam.

Inicialmente, irá ser feito um resumo sobre as LP, no Capítulo 2. No Capítulo 3, tratam-se dados, retirados de de uma base de dados credenciada, sobre o número de vítimas

resultantes de guerras e actos de terrorismo e faz-se a interpretação e o ajustamento dos mesmos usando LP. Finalmente, no último capítulo, enumeram-se as conclusões da pesquisa, onde se resumem os resultados obtidos e se apresenta trabalho futuro.

1.1. CONTEXTUALIZAÇÃO

Os eventos catastróficos naturais ou os causados pelo Homem têm características comuns, nomeadamente a distribuição estatística que melhora os aproxima. Exemplos de eventos causados pelo Homem são as guerras, actos de terrorismo e revoltas. Na natureza, os fenómenos mais estudados são os terremotos, deslizamentos de terra, erupções, cheias, entre outros [28].

As guerras têm tido um papel preponderante no decurso da história da Humanidade. Até ao século XVII, os estudos referentes a esta temática eram feitos com dificuldade. No entanto, era já possível identificar um padrão subjacente a estes eventos, nomeadamente, sabia-se que as guerras mais severas tinham ocorrências mais esporádicas, comparativamente a guerras com menor índice de mortalidade. Era, todavia, necessário incorporar este comportamento numa expressão mensurável [20].

Investigadores como M. Small e J.D. Singer [27] estudaram uma série de dados respeitantes a guerras. A explicação da relação entre intensidade das guerras e a sua frequência foi conseguida aproximando os dados a uma distribuição LP, sendo que quantos mais dados forem utilizados maior é a adequação a esta distribuição.

Outros autores como Bohorquez et al. [4] estudaram as relações entre o terrorismo global, guerras e ecologia. O seu objectivo principal era o de mostrar quantitativamente estas relações. Os autores partiram do pressuposto de que as organizações por detrás das guerras eram grupos organizados que evoluíam dinamicamente ao longo do tempo, com escolhas conscientes. As conclusões retiradas deste estudo foram: as guerras analisadas aproximavam-se à distribuição LP, com diferentes valores dos parâmetros estudados. Em particular, as guerras da Colômbia, Iraque, Perú e Afeganistão tinham valores diferentes dos parâmetros comparativamente às guerras Civis Espanhola e dos E.U.A.

No decorrer desta tese será estudada a aplicação das LP a fenómenos relacionados com guerras e actos de terrorismo, a nível internacional, regional e por tipologia.

1.2. OBJECTIVOS

O objetivo principal deste trabalho consiste no estudo das LP e na sua aplicação a guerras e actos de terrorismo.

Os pontos fundamentais tratados nesta tese são:

- Explicação das LP e dos seus parâmetros;
- Estado da arte das LP e suas aplicações mais comuns;
- Aplicação das LP a fenómenos relacionados com guerras e actos de terrorismo.

1.3. CALENDARIZAÇÃO

A calendarização das várias fases de desenvolvimento do trabalho é apresentada na Tabela 1. As tarefas desenvolvidas consistiram no estudo de bibliografia, que teve como base artigos científicos na área; na obtenção e análise de informação existente em bases de dados públicas confiáveis sobre guerras e actos de terrorismo. Por fim, elaborou-se o relatório final. O estudo de bibliografia foi a tarefa que se prolongou mais no tempo, de maneira a poder relembrar conceitos antigos dados em cadeiras relativas a estatística e à assimilação de novos conceitos sobre leis de potência. A duração do relatório final deve-se sobretudo à componente teórica da tese.

Tabela 1 Calendarização do Projeto

		Out.				Nov.				Dez.				Jan.				Fev.				Mar.				Abr.				Mai.				Jun.			
ETAPAS	SEMANAS	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Recolha da bibliografia	10 S																																				
Estudo da bibliografia	12 S																																				
Implementação	10S																																				
Análise dos dados	10S																																				
Elaboração do relatório final	7S																																				

1.4. ORGANIZAÇÃO DO RELATÓRIO

No Capítulo 1, é feita uma introdução às LP. No capítulo seguinte, é apresentada uma descrição mais detalhada das LP e as suas principais aplicações, nomeadamente, as relacionadas com guerras e actos de terrorismo. No Capítulo 3, aplicam-se as LP a dados relativos a guerras e actos de terrorismo, desenhando os gráficos das distribuições cumulativas dos fenómenos estudados. Por último, no Capítulo 4 são reunidas as principais conclusões do trabalho desenvolvido e indicam-se direcções de pesquisa futura.

2. LEIS DE POTÊNCIA

A Humanidade tem enfrentado inúmeros conflitos, muitos deles catastróficos, pelo nível de destruição causado e, muito principalmente, pelo número de vítimas que fizeram. Surge, assim, uma necessidade urgente de quantificar estes dados, usando uma ferramenta capaz de explicar os padrões de comportamento destes fenómenos e possíveis futuras tendências. Uma dessas ferramentas é a distribuição estatística chamada Lei de Potência (LP). Neste capítulo explica-se a LP, como são calculados os seus parâmetros e como vários autores utilizaram esta ferramenta na análise do número de vítimas de guerras, guerrilhas e actos de terrorismo.

2.1. LEI DE POTÊNCIA

A função de probabilidade de uma variável aleatória discreta, X , que toma apenas valores inteiros e que segue uma distribuição LP é dada por [13]:

$$p(x) = P(X = x) = Cx^{-\alpha} \quad (1)$$

onde:

- $p(x)$: probabilidade de uma variável tomar o valor x ;

- α : expoente da distribuição (tipicamente $\alpha > 1$);
- x : valor que a variável aleatória X toma;
- C : constante.

Esta distribuição diverge em zero, pelo que consideramos que existe um valor x_{min} correspondente ao valor mínimo de x em que o comportamento da LP é observado. O cálculo da constante C resulta da normalização da função de probabilidade, isto é, verifica-se que:

$$1 = \sum_{x_{min}}^{\infty} p(x) = C \sum_{x_{min}}^{\infty} x^{-\alpha} \quad (2)$$

obtendo-se [8]:

$$p(x) = \frac{x^{-\alpha}}{\zeta(\alpha, x_{min})} \quad (3)$$

onde $\zeta(\alpha, x_{min}) = \sum_0^{\infty} (n + x_{min})^{-\alpha}$ é a função zeta de Hurwitz.

A função densidade de probabilidade (fdp) de uma variável aleatória contínua, X , que segue uma distribuição LP é dada por [8]:

$$f(x) = Cx^{-\alpha} \quad (4)$$

Esta função densidade também diverge quando x tende para zero, pelo que existe um valor x_{min} , como acima descrito. Assumindo que $\alpha > 1$ e calculando C por normalização da fdp vem:

$$1 = \int_{x_{min}}^{\infty} f(x)dx = C \int_{x_{min}}^{\infty} x^{-\alpha}dx = \frac{C}{1-\alpha} [x^{-\alpha+1}]_{x_{min}}^{\infty} \quad (5)$$

obtendo-se:

$$C = (\alpha - 1)x_{min}^{\alpha-1} \quad (6)$$

Logo, a fdp de uma variável aleatória contínua X que segue uma distribuição LP é dada por:

$$f(x) = \frac{\alpha-1}{x_{min}} \left(\frac{x}{x_{min}} \right)^{-\alpha} \quad (7)$$

A função de distribuição acumulada complementar é a função que retorna a probabilidade de uma variável aleatória X tomar valores maiores ou iguais a x . A função de distribuição devolve o valor da probabilidade de uma variável X tomar valores menores ou iguais a x . No caso de X ser uma variável aleatória discreta com uma distribuição LP, a função de distribuição acumulada complementar é a seguinte [8]:

$$F(x) = F(X \geq x) = \frac{\zeta(\alpha, x)}{\zeta(\alpha, x_{min})} \quad (8)$$

Se X é uma variável aleatória contínua, então, a função de distribuição complementar é dada por:

$$F(x) = \left(\frac{x}{x_{min}} \right)^{-\alpha+1} \quad (9)$$

Como se pode observar, as fórmulas são mais simples quando X é uma variável aleatória contínua. Em muitas aplicações é comum aproximar dados discretos por uma função LP contínua. Isto pode ser feito considerando que os dados são arredondamentos, para o inteiro mais próximo, de valores gerados por uma LP contínua. No que se segue, consideramos X uma variável aleatória contínua.

2.2. MÉDIA E VARIÂNCIA DA LP

A média e variância amostrais de uma distribuição LP podem ser obtidas com base no método dos momentos (MM). O MM é um método que tem como base o procedimento seguinte [26]:

- 1º passo: Igualar os momentos da população (sendo estas funções de parâmetros desconhecidos) aos momentos amostrais;
- 2º passo: Resolver as equações em função dos parâmetros desconhecidos.

Passamos a apresentar as fórmulas gerais dos primeiros momentos da população e da amostra:

$$\text{Primeiro momento (média) da população: } E(x) = \mu \quad (10)$$

$$\text{Primeiro momento (média) da amostra: } M_1 = \bar{X} \quad (11)$$

Seguindo o primeiro passo do procedimento tem-se a igualdade $\hat{\mu} = \bar{X}$, isto é, conclui-se que a média amostral é o estimador da média da população.

Aplicando o passo inicial à distribuição LP, é necessário encontrar o primeiro momento da amostra, sendo esta dada pela expressão seguinte [18]:

$$\bar{x} = \int_{x_{min}}^{\infty} x f(x) dx = \frac{C}{2-\alpha} [x^{-\alpha+2}]_{x_{min}}^{\infty} \quad (12)$$

Da equação acima, pode-se concluir que, para valores de **α inferiores a 2**, a função diverge, isto é, não tem média finita. Esta conclusão é puramente teórica, pois seria preciso, na prática, um número infinito de amostras para mostrar que a média é infinita. Para α superior a 2, o primeiro momento, que corresponde à media amostral, é dado pela igualdade:

$$\langle x \rangle = \bar{x} = \frac{\alpha-1}{\alpha-2} x_{min} \quad (13)$$

É possível calcular os momentos seguintes, sendo o segundo momento, representando a média quadrática da distribuição, dado pela expressão:

$$\langle x^2 \rangle = \frac{C}{3-\alpha} [x^{-\alpha+3}]_{x_{min}}^{\infty} \quad (14)$$

Da equação acima, pode-se concluir que, para valores de α inferiores a 3, a função diverge, isto é, não tem média quadrática, e, por consequência, não tem variância nem desvio-padrão finitos. Para α superior a 3, o segundo momento é o seguinte:

$$\langle x^2 \rangle = \frac{\alpha-1}{\alpha-3} x_{min}^2 \quad (15)$$

Todos os momentos seguintes podem ser obtidos através da expressão:

$$\langle x^m \rangle = \frac{\alpha-1}{\alpha-1-m} x_{min}^m \quad (16)$$

A expressão (16) só é válida para valores de m inferiores a $\alpha - 1$, uma vez que se forem superiores a média irá divergir.

2.3. ESTIMAÇÃO DE PARÂMETROS DA LEI DE POTÊNCIA

A estimação dos parâmetros C e α é muito importante, para se conseguir uma boa aproximação de um conjunto de dados a uma distribuição LP. Neste capítulo, vamos explicar como se obtém cada um destes parâmetros, usando o método da máxima verossimilhança e o método dos mínimos quadrados.

Os valores de C e de α apresentados acima correspondem aos valores reais dos parâmetros, tendo como base dados de determinada população. Considerando que é muito difícil obter os dados da totalidade da população, o mais habitual é utilizar-se uma amostra da população e estimar-se os valores aproximados dos parâmetros, dados por \hat{C} e $\hat{\alpha}$.

Por fim são apresentados alguns métodos para estimar o valor de x_{min} .

- Método da máxima verossimilhança (MMV)

O MMV é um método que permite estimar os valores dos parâmetros de uma população, a partir de dados amostrais. Escolhe estimativas que maximizam a probabilidade de obter aquela amostra [22]. No princípio de verosimilhança estatística são os dados que decidem quais os modelos apropriados e não o inverso [12].

Sendo a fdp uma função calculada através de uma amostra aleatória (X_1, X_2, \dots, X_n) , a função verosimilhança de θ é representada da seguinte forma [16]:

$$L(\theta|x) = \prod_{i=1}^n f(x_i|\theta) \quad (17)$$

Quanto ao estimador de MV, regra geral segue a inequação:

$$L(\tilde{\theta}|x) \geq L(\theta|x), \forall \theta \in B \quad (18)$$

Isto é, $\tilde{\theta}$ é uma estimativa de θ se este maximizar os valores da função verosimilhança dada em (17).

O segundo passo consiste em calcular o logaritmo da função verosimilhança, obtendo assim a transformada logarítmica (mais utilizada para efeitos práticos):

$$\ln L(\theta|x) = \sum_{i=1}^n \ln f(x_i|\theta) \quad (19)$$

O último passo consiste em calcular o máximo da função (19), de modo a obter o estimador segundo o MMV:

$$\frac{\partial \ln L(\theta|x)}{\partial \theta} = 0 \quad (20)$$

Vamos agora aplicar o método da função de máxima verosimilhança à distribuição LP. Consideramos a fdp dada na expressão (7) e um conjunto de n observações $x_i \geq x_{min}$. Vamos calcular o valor de α que gerou este conjunto de dados a partir da distribuição LP. A probabilidade dos dados terem sido gerados pela LP é proporcional a [18]:

$$L(\alpha) = \prod_{i=1}^n f(x_i) = \prod_{i=1}^n \frac{\alpha-1}{x_{min}} \left(\frac{x_i}{x_{min}} \right)^{-\alpha} \quad (21)$$

Calculando o logaritmo da função (21) obtém-se:

$$\begin{aligned} \ln(L(\alpha)) &= \ln f(x|\alpha) = \sum_{i=1}^n \left[\ln(\alpha-1) - \ln x_{min} - \alpha \ln \frac{x_i}{x_{min}} \right] = \\ &= n \ln(\alpha-1) - n \ln x_{min} - \alpha \sum_{i=1}^n \ln \frac{x_i}{x_{min}} \end{aligned} \quad (22)$$

O máximo da função (22) é o zero da sua primeira derivada. Resolvendo a expressão obtida em ordem a α , obtém-se o estimador de MV $\hat{\alpha}$, dado por:

$$\frac{\partial \ln L(\alpha)}{\partial \alpha} = 0 \Leftrightarrow \frac{n}{\alpha-1} - \sum_{i=1}^n \ln \frac{x_i}{x_{min}} = 0 \Leftrightarrow \hat{\alpha} = 1 + n \left[\sum_{i=1}^n \ln \frac{x_i}{x_{min}} \right]^{-1} \quad (23)$$

- Método dos mínimos quadrados (MMQ)

O método dos mínimos quadrados permite estimar os valores dos parâmetros a e b de uma recta do tipo $Y = a + bX$, que aproxima os dados de uma população, tendo como objectivo a minimização da soma do quadrado dos resíduos. Os resíduos consistem nas diferenças entre os valores estimados e os valores reais.

O método dos mínimos quadrados minimiza a soma do quadrado dos resíduos [10]:

$$S = \sum_{i=1}^n (Y_i - a - bX_i)^2 \quad (24)$$

Derivando a expressão (24) e igualando a zero é possível encontrar os estimadores \hat{a} e \hat{b} dos parâmetros a e b da recta, obtendo-se [29]:

$$\hat{b} = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \quad (25)$$

$$\hat{a} = \frac{\sum_{i=1}^n y_i - b \sum_{i=1}^n x_i}{n} \quad (26)$$

Para calcular os parâmetros de uma distribuição LP pelo método dos mínimos quadrados, segue-se o seguinte procedimento. A função de distribuição complementar (9) pode ser linearizada, aplicando a função logaritmo a ambos os lados da igualdade. Obtém-se:

$$\ln F(x) = (\alpha - 1) \ln x_{min} + (1 - \alpha) \ln x \quad (27)$$

com:

$$Y = \ln F(x); a = (\alpha - 1) \ln x_{min}; b = 1 - \alpha; X = \ln x \quad (28)$$

Basta aplicar agora as fórmulas (27) e (28) e a partir daí aplicar as fórmulas abaixo para retirar os estimadores $\hat{\alpha}$ e \hat{C} .

$$\hat{\alpha} = 1 - \hat{b}; \quad \hat{C} = (\hat{\alpha} - 1)e^{\hat{a}} \quad (29)$$

O cálculo do limite inferior x_{min} a partir do qual uma dada distribuição deixa de ter comportamento de uma LP pode constituir um desafio. É frequente existirem dados cujo comportamento não se adequa a uma LP na “cauda” da distribuição [8]. A estimativa de x_{min} tem de ser um valor intermédio entre uma estimativa sobrevalorizada (onde se ignoram dados relevantes e se aumenta o erro estatístico), e uma estimativa subvalorizada (onde estarão a ser incluídos dados que não acrescentam valor à explicação da distribuição).

O método mais usual e mais simples na determinação de x_{min} é através da visualização num gráfico, em dupla escala logarítmica, do ponto onde os dados deixam de estar alinhados numa recta e começam a ter um comportamento exponencial. A esse conjunto de dados não se aplica uma LP. Existe, no entanto, outro método de cálculo deste parâmetro, mais complexo e calculado através do Critério de Informação Bayesiano (CIB) onde a maximização do CIB, através da maximização da verosimilhança marginal, gerará uma estimativa de \hat{x}_{min} ; A expressão do CIB esta sujeita à seguinte aproximação [8]:

$$\ln Pr(x|x_{min}) \cong L - \frac{1}{2}(x_{min} + 1) \ln n \quad (30)$$

2.3.1 ESTIMAÇÃO DE PARÂMETROS: NOTAS FINAIS

Bauke [2] fez um estudo onde compara estimadores de LPs discretas e chegou à conclusão de que o MMQ não pode ser utilizado. Os dados gerados por uma distribuição LP não seguem os pressupostos gerais da regressão linear sendo assumido, por exemplo, que os dados X_i não têm erro e que os dados Y_i podem conter erro. Estes últimos erros são considerados variáveis aleatórias independentes e identicamente distribuídas com média zero. Em particular, o desvio-padrão de Y_i é independente de X_i . Bauke [2] apresenta como alternativa o MMV, um método que segundo ele é assintoticamente eficiente, para aplicar em dados com o comportamento de uma LP.

2.4. TESTES PARAMÉTRICOS : GOODNESS-OF-FIT

Na secção anterior foram referidos métodos que permitem estimar/ajustar os parâmetros de uma distribuição LP. Todavia, nada foi dito relativamente à qualidade dessa estimação/ajuste. O teste de Kolmogorov-Smirnoff (KS) [8] é um teste não paramétrico que pode ser usado para comparar uma amostra com uma distribuição de probabilidade de referência (teste KS para uma amostra). A estatística de KS quantifica a distância entre a função de distribuição empírica da amostra $S(x)$ e a função de distribuição da distribuição de referência $P(x)$, dada por:

$$D = \max_{x \geq x_{min}} |S(x) - P(x)| \quad (31)$$

A hipótese nula afirma que a amostra é retirada da distribuição de referência. As distribuições consideradas na hipótese nula têm de ser contínuas.

Ao contrário da maioria dos testes de hipótese, a não rejeição da hipótese nula é desejável, uma vez que significa que a distribuição não pode ser excluída para a explicação dos dados.

A aplicação prática deste teste para uma distribuição LP segue o seguinte procedimento:

1. Estimar α e x_{min} através de um método de estimação de parâmetros (MMV, MMQ);
2. Realizar o teste KS para gerar um valor de estatística D para a distribuição LP com parâmetros calculados em 1.
3. Gerar um grande número de bases de dados com os parâmetros calculados em 1. de modo a obter diferentes estatísticas D , usando o Método da Transformação, por exemplo [8];
4. Calcular o p -value como fracção das estatísticas dos testes KS gerados a partir das diferentes bases de dados, cujo valor ultrapassa a estatística KS da distribuição LP de 2.;
5. Se o p -value for suficiente baixo, então a distribuição LP pode ser excluída como um bom ajuste para os dados.

A comparação com outros modelos pode ser feita seguindo o procedimento acima, desde que se adapte os métodos à distribuição que se vai comparar (por exemplo distribuição exponencial, lognormal) extraíndo daí as estatísticas KS e, consequentemente, o p -value respectivo. Se o p -value for muito pequeno podemos excluir essa distribuição, sendo que quantas mais distribuições compararmos e excluirmos versus uma LP, maior é a probabilidade de os dados se aproximarem a esta distribuição. Claro que, o número de comparações da LP com distribuições semelhantes seria infinito. Por outro lado, é sempre possível encontrar uma distribuição que se ajusta melhor aos dados do que uma LP, se definirmos uma família de curvas com um número suficientemente grande de parâmetros. O ajuste de uma distribuição estatística deve assim incluir técnicas estatísticas e um pouco de intuição física.

2.5. APLICAÇÃO ÀS GUERRAS E ACTOS DE TERRORISMO

Richardson [24] foi um dos primeiros a estudar a aplicação das LPs às guerras e actos de terrorismo, em 1948. Este autor estudou dados relativos à violência entre 1820 e 1945, introduzindo o conceito de divisão por categorias, dependendo da severidade das guerras.

Chegou à conclusão de que as categorias onde se geravam mais casualidades/vítimas eram aquelas que tinham menor frequência. Os investigadores M. Small e J.D. Singer [27] recolheram dados relativos a 119 guerras entre 1816 e 1980 e classificaram-nas de acordo com a sua intensidade. A intensidade foi definida como a relação entre o número de mortos e o número de habitantes do país, sendo expressa por x mortos/ 10000 habitantes. Os autores observaram que as guerras com menos vítimas (até 10 mortos por 10000 habitantes) eram muito frequentes, sendo que as guerras mais violentas (intensidades acima dos 100) eram reduzidas. A guerra mais violenta foi a que ocorreu entre Paraguai e Bolívia, no período de 1932 a 1935, com uma intensidade de 382,4.

Roberts e Turcotte [25], em 1998, estudaram o número de vítimas de 119 guerras entre 1945 e 1973. Chegaram à conclusão de que a relação entre a intensidade (medida em termos do número de vítimas) e a frequência das guerras era bem aproximada por uma LP. Os parâmetros da LP eram semelhantes aos de Richardson [24].

Em 2002, Cederman [6] seguiu o trabalho de Richardson, introduzindo o conceito de modelo baseado em agentes. Neste tipo de modelos, considera-se que os agentes envolvidos são organizados e tomam decisões lógicas e conscientes, de modo a poder prever tendências de guerras, que se alastram geograficamente ou se prolongam no tempo. Cederman utilizou dados retirados do projecto Correlates of War (COW) [11] e concluiu que os dados seguiam uma distribuição LP. Em 2006, Bogen and Jones [3] interpretaram os dados relativos a 27000 incidentes entre 1968 e 2004 recolhidos da Corporação RAND e do Instituto de Prevenção do Terrorismo da Cidade de Oklahoma. Classificaram os dados em termos de severidade e concluíram que a sua distribuição se aproximava de uma LP. Usaram esta LP para fazer previsões, até 2080, da mortalidade causada pelas guerras. Com isto, mostraram que a LP não serve apenas para interpretação, mas também como ferramenta de previsão destes fenómenos, de modo a controlar e atenuar as suas consequências.

Em 2006, Johnson et al [14] provaram que as guerras do Iraque, Afeganistão, Colômbia e países não pertencentes ao G7 seguiam uma distribuição semelhante à LP. Os autores demonstraram que os padrões vistos na evolução das diferentes guerras eram semelhantes, independentemente das motivações por detrás destas.

Clauset et al [9], em 2007, focaram-se no binómio frequência/severidade das guerras, recolhendo dados de guerras entre 1968 e 2007. Os gráficos log-log obtidos a partir dos

dados mostraram que a generalidade das guerras seguia uma distribuição LP e permitiram evidenciar uma tendência periódica das guerras com mais vítimas, no caso dos dados analisados em períodos de 13 anos.

3. APLICAÇÃO A DADOS REAIS DE GUERRAS E ACTOS DE TERRORISMO

Neste capítulo aplicam-se as LP a dados reais de guerras e actos de terrorismo. Estuda-se a distribuição cumulativa das variáveis associadas a cada fenómeno. O tratamento dos dados é feito usando o procedimento seguinte:

1. Recolha de dados referentes ao tema em estudo;
2. Tratamento dos dados, agrupando em categorias as guerras, sendo o critério de classificação a frequência do nº de baixas;
3. Ordenar a frequência do nº de baixas por ordem decrescente e efectuar a sua classificação de 1 até n, onde 1 representa a categoria de maior frequência (representação gráfica antes da normalização, ver Figura 1);
4. Adequar escalas logarítmicas a cada um dos eixos (ao eixo dos xx representado pela frequência e ao eixo dos yy representado pela classificação);

5. Normalizar os dados através da divisão de todos os valores no eixo dos xx pelo máximo que tomam, repetindo o processo para o eixo dos yy;
6. Representar a recta aproximada aos dados do gráfico, sendo a reta calculada minimizando a soma do quadrado dos resíduos (ver Figura 2);
7. Apresentação da recta e do coeficiente de determinação, R^2 , respectivo. Valores próximos de 1 do coeficiente R^2 indicam que os dados são bem ajustados pela recta, isto é, que a distribuição estatística que melhor se lhes adequa é uma LP.

3.1. NÚMERO DE MORTOS EM GUERRAS A NÍVEL MUNDIAL

Os dados utilizados neste estudo provêm da Global Terrorism Database: <http://www.start.umd.edu/gtd/>, retirados em Janeiro de 2014. Os dados estudados englobam informação sobre o número de vítimas em 3693 guerras e actos de terrorismo, a nível mundial, no período de 1972 a 2012. Os dados foram submetidos ao procedimento descrito acima. O output produzido encontra-se sobre forma de gráfico log-log, onde os eixos representam frequência/classificação.

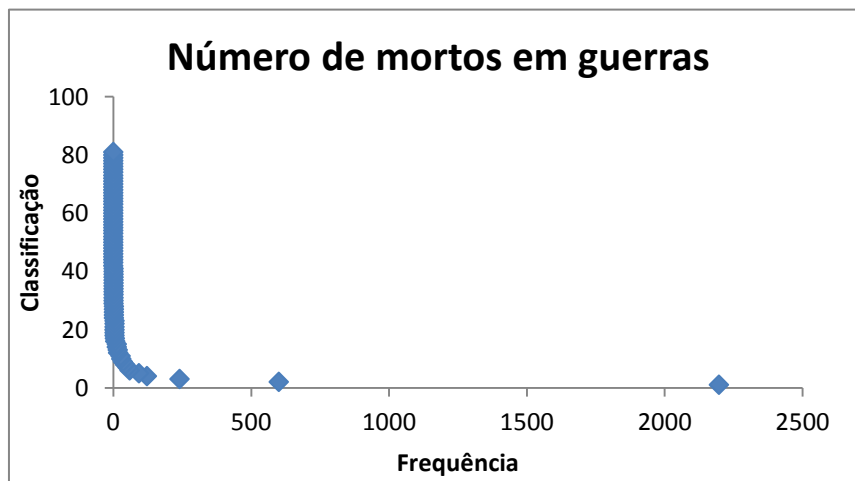


Figura 1 Gráfico classificação/frequência antes da normalização

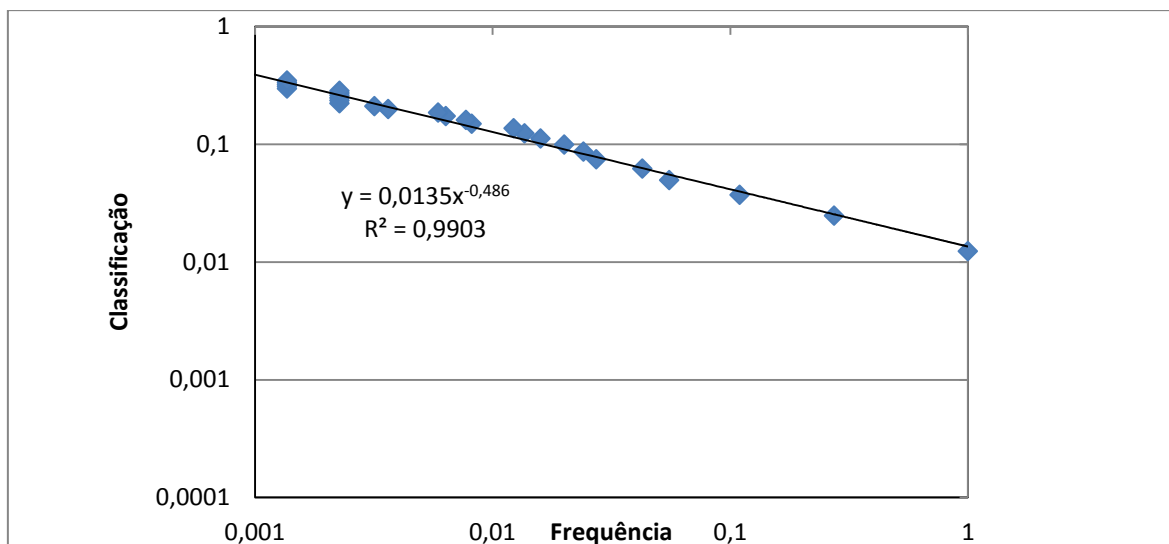


Figura 2 Gráfico classificação/frequência após normalização

As Figuras 1 e 2 representam número de mortos em guerras e actos de terrorismo, referentes à totalidade da base de dados analisada. O nº de mortos encontra-se entre os 0 e os 1382.

Uma vez analisada a base de dados na sua totalidade e tendo sido aproximada a sua distribuição, vamos passar a analisar os dados impondo certas condições, como sendo a distribuição regional, a tipologia das guerras e o período histórico.

3.2. NÚMERO DE MORTOS EM GUERRAS POR REGIÃO

Os dados analisados referem-se a guerras e actos de terrorismo na Europa, Europa de Leste, América do Norte, América Central e Caraíbas, América do Sul, Austrália e Oceânia, África e Ásia. As regiões serão reduzidas para seis neste estudo, sendo Europa, América do Norte, América Central e Sul, África, Ásia e Austrália.

3.2.1 América do Norte

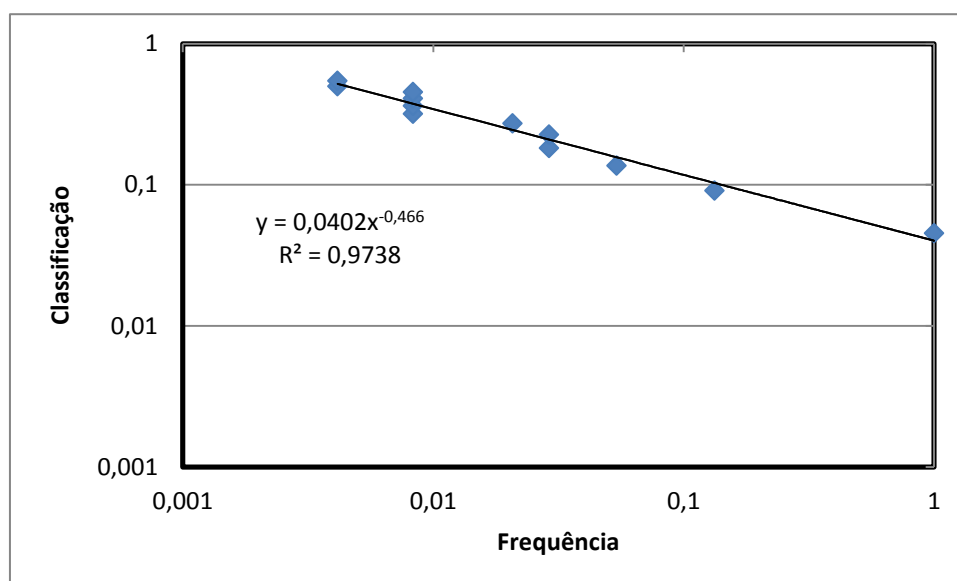


Figura 3 Gráfico classificação/frequência após normalização: América do Norte

Analísámos e tratámos, segundo o procedimento descrito no início deste capítulo, os dados relativos a guerras na América do Norte. O gráfico na dupla escala logarítmica, permite-nos observar uma boa adequação à distribuição LP (Figura 3). O coeficiente de determinação é muito próximo de 1.

3.2.2 América Central e Sul

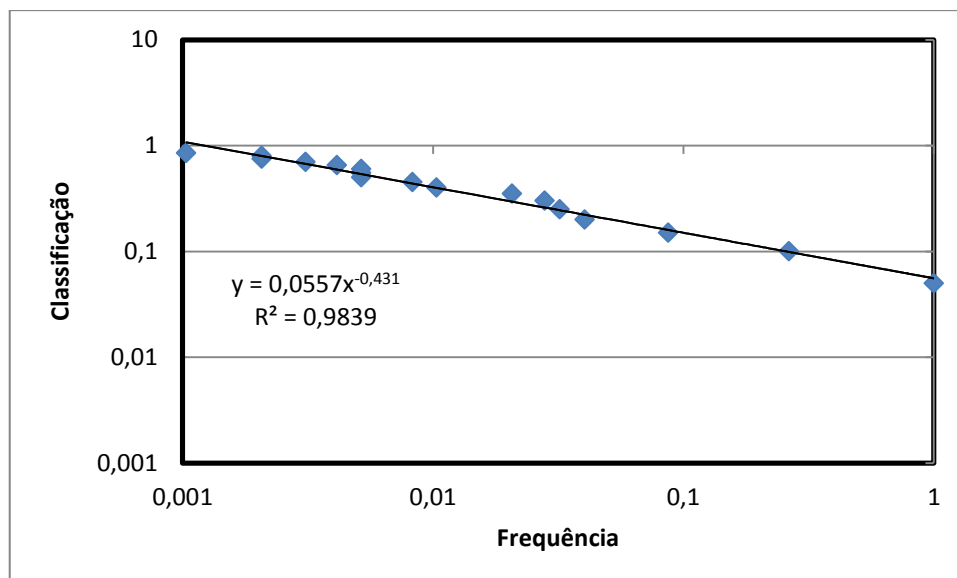


Figura 4 Gráfico classificação/frequência após normalização: América Central e Sul

Analísamos e tratámos, segundo o procedimento descrito no início deste capítulo, os dados relativos a guerras na América Central e do Sul. O gráfico na dupla escala logarítmica, permite-nos observar uma boa adequação à distribuição LP (Figura 4). O coeficiente de determinação é muito próximo de 1.

3.2.3 Europa

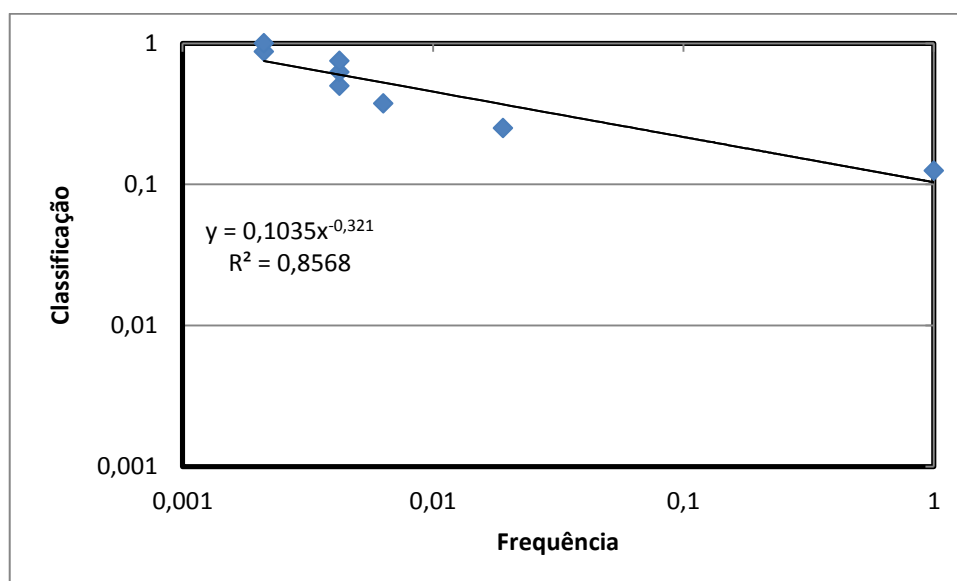


Figura 5 Gráfico classificação/frequência após normalização: Europa

Analisámos e tratámos, segundo o procedimento descrito no início deste capítulo, os dados relativos a guerras e actos de terrorismo na Europa. O gráfico na dupla escala logarítmica, permite-nos observar um fraco ajustamento à distribuição LP (Figura 5). O coeficiente de determinação é de apenas 0.85, distante do valor óptimo 1, para casos de um excelente ajustamento. Este fenómeno pode ser explicado pelo insuficiente número de dados disponíveis para tratamento. Os outliers têm, assim, uma influência importante neste ajustamento.

3.2.4 Ásia

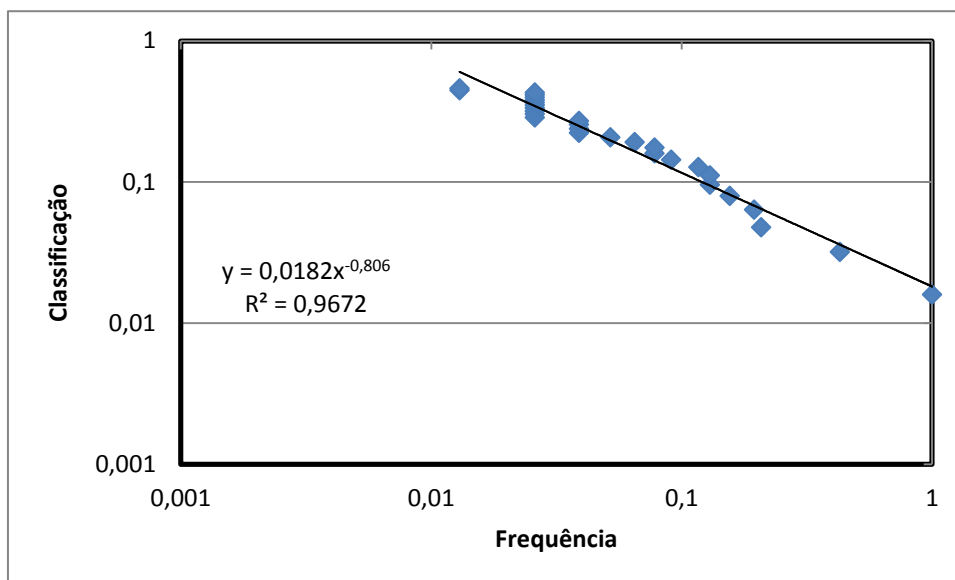


Figura 6 Gráfico classificação/frequência após normalização: Ásia

Os dados tratados e analisados relativos à Ásia são bem aproximados por uma LP (Figura 6). O coeficiente de determinação tem um valor próximo de 1. Neste caso, o declive da recta que aproxima os dados é mais acentuado do que nos restantes casos estudados. Isto pode ser explicado pelo facto de, nesta região, existirem guerras com maior gravidade e maior frequência do que nas restantes regiões.

3.2.5 Austrália

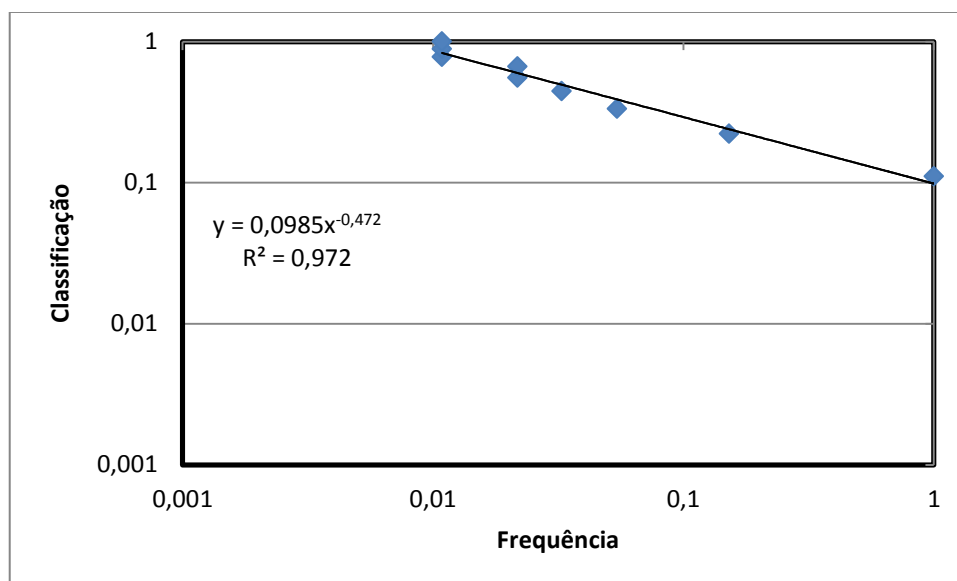


Figura 7 Gráfico classificação/frequência após normalização: Austrália

Os dados analisados e tratados, referentes à Austrália, revelam uma boa aproximação a uma distribuição LP. O coeficiente de determinação é próximo de 1 (Figura 7).

3.2.6 África

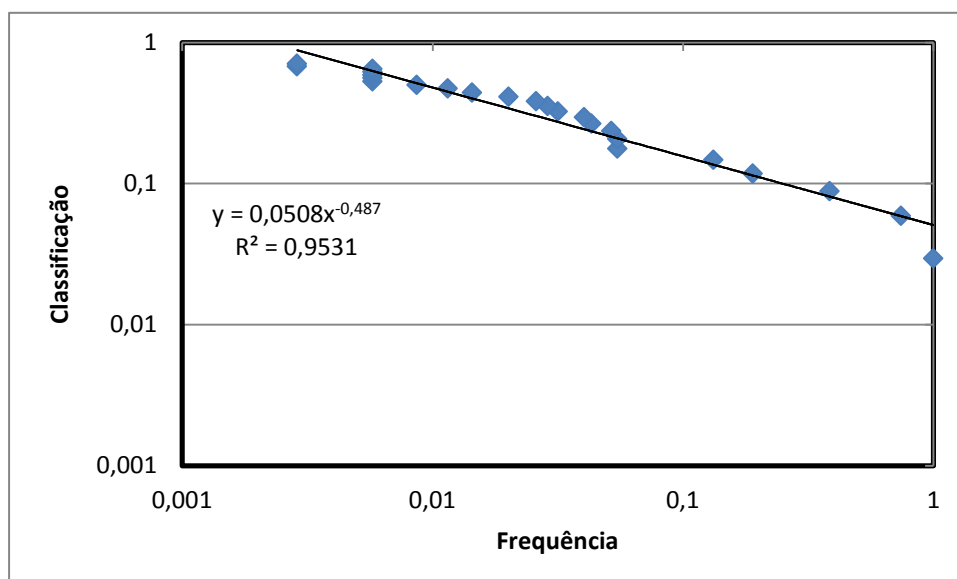


Figura 8 Gráfico classificação/frequência após normalização: África

Por fim, analisámos e tratámos dados do continente africano. Tal como a maioria das regiões analisadas, os dados adequam-se a uma distribuição de LP. O coeficiente de determinação é próximo de 1 (Figura 8).

3.2.7 Discussão dos resultados a nível regional

Na Figura 9 representa-se a relação entre os parâmetros das LP para cada região.

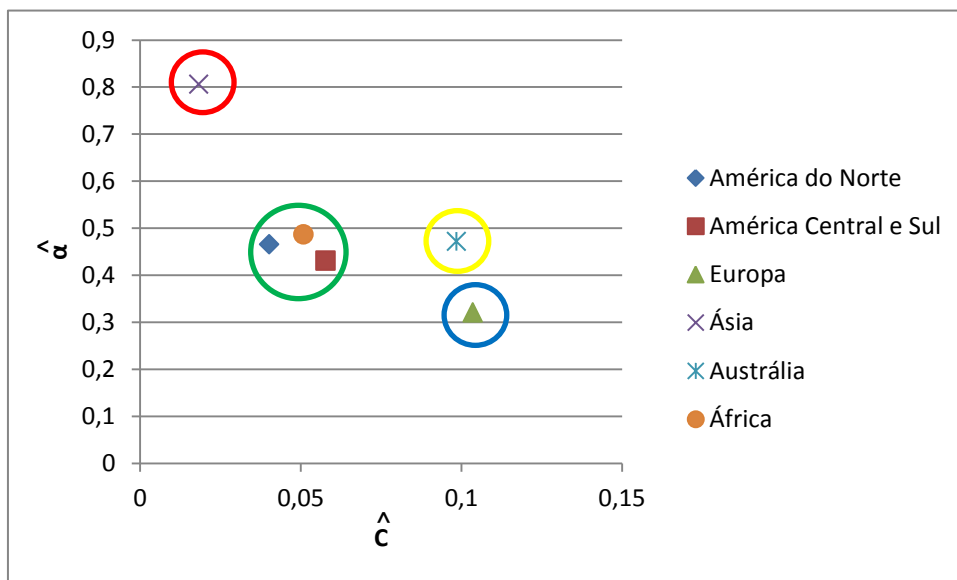


Figura 9 Representação dos parâmetros \hat{C} e \hat{a} das guerras por região

Os valores calculados de \hat{C} estão situados no intervalo $0.0182 \leq \hat{C} \leq 0.1035$, onde os valores mínimos e máximos correspondem à Ásia e Europa, respectivamente. No que respeita ao parâmetro \hat{a} os valores estimados encontram-se entre $0.321 \leq \hat{a} \leq 0.806$, sendo os valores mínimos e máximos associados novamente à Europa e Ásia.

A representação gráfica permite-nos identificar quatro zonas distintas. Uma zona (delimitada a vermelho) com um maior número de mortos em guerras, a Ásia, com valores elevados do parâmetro \hat{a} e valores reduzidos do parâmetro \hat{C} . A segunda zona (a verde) do

gráfico engloba os parâmetros das LP referentes à América (Norte, Central, Sul) e África. Os dados das guerras dessas regiões têm padrões de comportamento muito semelhantes, originando LP com valores de parâmetros muito próximos. A terceira zona (a amarelo) refere-se à Austrália, com um comportamento diferente das restantes regiões, uma vez que apresenta um parâmetro \hat{C} maior à excepção da Europa. Por último a quarta zona (a azul) refere-se à Europa, caracterizada por um número menor de mortos, com valores reduzidos do parâmetro $\hat{\alpha}$ e valores elevados do parâmetro \hat{C} .

Em termos regionais as conclusões tiradas da observação dos diferentes casos analisados, são as seguintes:

- a maioria das regiões analisadas possuem dados que se adequam a uma distribuição LP, com valores do coeficiente de determinação R^2 superiores a 90%, à excepção da Europa;
- os resultados obtidos para a Europa podem ser explicados pelo pequeno número de classificações disponibilizado, face a outras regiões. Neste caso, os outliers têm grande influência aquando do ajustamento de uma distribuição LP;
- a região da Ásia teve guerras mais severas com maior frequência de vítimas, contrariamente a zonas do globo como a Europa, com frequência e nº de mortos menores face às restantes regiões. O acto de terrorismo com maior número de baixas ocorreu na América do Norte (11 de Setembro de 2001). Europa e Austrália são continentes onde as guerras não aconteceram com tanta frequência nem tiveram tantas baixas, no período estudado. Valores elevados de \hat{C} estão associados a menor severidade, o mesmo acontece com valores elevados de $\hat{\alpha}$.

3.3. NÚMERO DE MORTOS EM GUERRAS POR TIPOLOGIA DE GUERRA

Para estudar a distribuição dos dados consoante a tipologia de guerra, os dados recolhidos e tratados foram divididos em duas grandes categorias: guerrilhas e guerras religiosas. As guerrilhas têm motivações de ordem económica/social, política, entre outras. Encontram-se incluídos nas guerrilhas atentados bombistas, sem nenhum grupo terrorista associado, guerras civis e ataques organizados de células terroristas a civis ou militares. As duas

grandes guerras mundiais não foram incluídas neste trabalho, pelo facto de não terem decorrido no horizonte temporal dos dados recolhidos. As guerras com motivações religiosas, em que os grupos terroristas, que as levam a cabo, são crentes das mais diversas religiões mundiais, têm como alvos preferenciais as instituições religiosas, tais como igrejas, templos, mesquitas, entre outros.

Iremos passar de seguida à análise de cada um destes tipos de guerra e retirar as conclusões mais pertinentes.

3.3.1 Guerrilhas

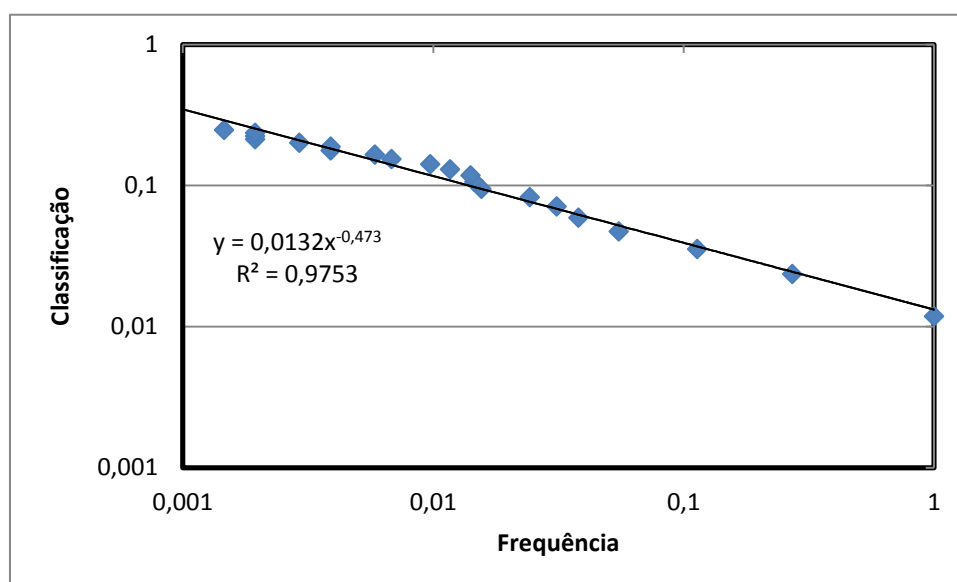


Figura 10 Gráfico classificação/frequência após normalização: Guerrilhas

Na Figura 10 podemos observar que o número de vítimas resultantes de actividades de guerrilha se aproxima bem de uma recta, num gráfico log-log. Os dados são bem aproximados por uma distribuição LP.

3.3.2 Guerras religiosas

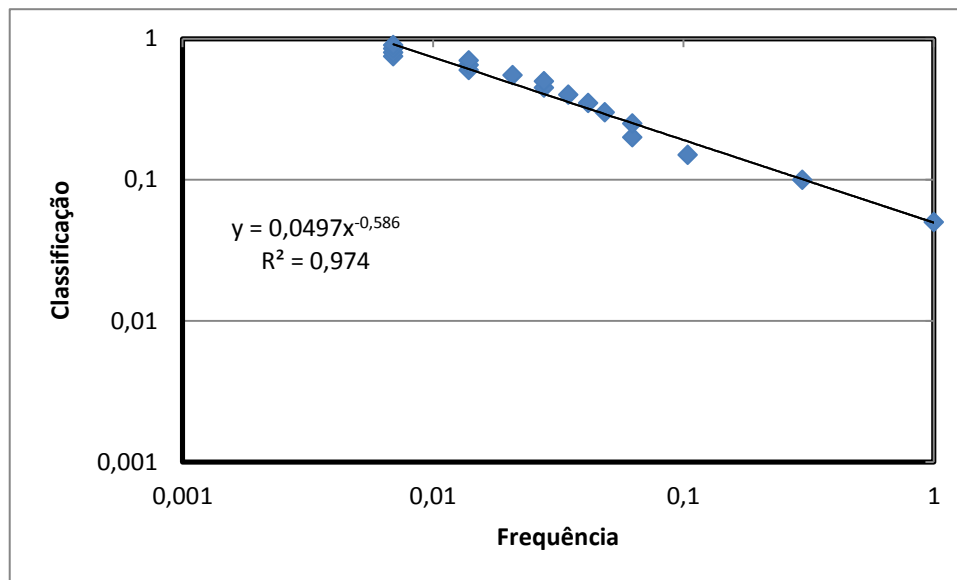


Figura 11 Gráfico classificação/frequência após normalização: Religiosas

As guerras de carácter religioso (Figura 11) têm habitualmente um número de vítimas baixo, situando-se entre as 10 e as 200. Os dados são bem aproximados, após normalização, por uma distribuição semelhante a uma LP.

Na Figura 12, podemos observar a relação entre os parâmetros das LP para cada tipo de guerra.

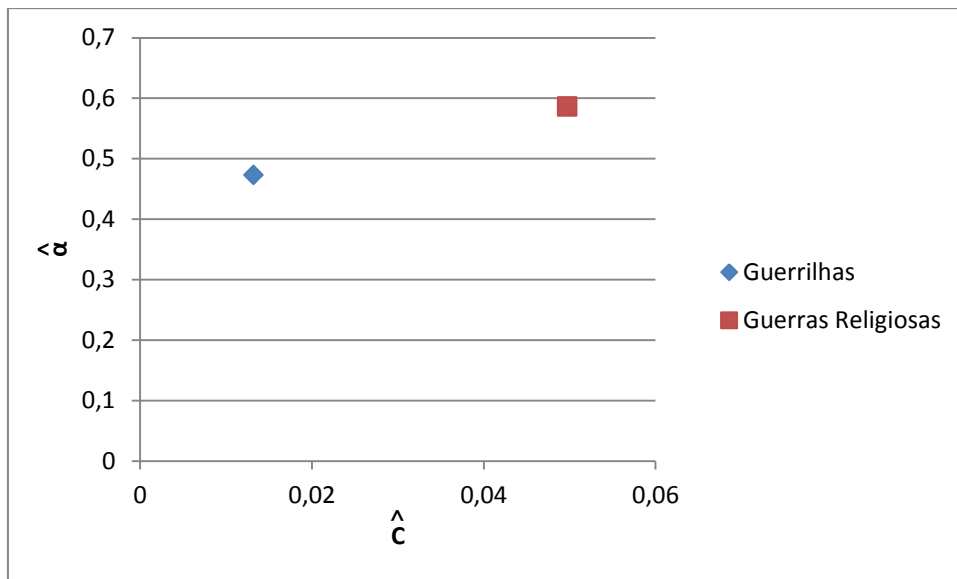


Figura 12 Representação dos parâmetros \hat{C} e $\hat{\alpha}$ por tipologia de guerra

Os valores calculados de \hat{C} e $\hat{\alpha}$ encontram-se nos intervalos $0.0132 \leq \hat{C} \leq 0.0497$ e $0.473 \leq \hat{\alpha} \leq 0.586$, respectivamente. O valor de \hat{C} varia significativamente, ao contrário de $\hat{\alpha}$. Há uma distinção bem acentuada entre os dois tipos de guerras. Em conclusão, a distribuição dos dados dos dois tipos de guerras adequa-se a uma LP, com um coeficiente de determinação da mesma ordem de grandeza. Salienta-se que existem mais guerras do tipo guerrilhas do que de carácter religioso, sendo estas mais severas do que as anteriores e estando-lhes associados valores mais baixos de \hat{C} .

3.4. NÚMERO DE MORTOS EM GUERRAS POR PERÍODO HISTÓRICO

Nesta seção iremos fazer o estudo do número de vítimas mortais em períodos de 10 anos (à excepção do período de 1972-1992). Pretendemos estudar a distribuição dos dados de guerras nestes períodos e verificar se se trata de uma LP.

3.4.1 Período de 1972-1992

O primeiro período em análise corresponde ao período de 1972 e 1992. Este é o único período que engloba dados relativos a 20 anos de guerras, devido ao número reduzido de observações disponibilizadas pela base de dados utilizada neste estudo.

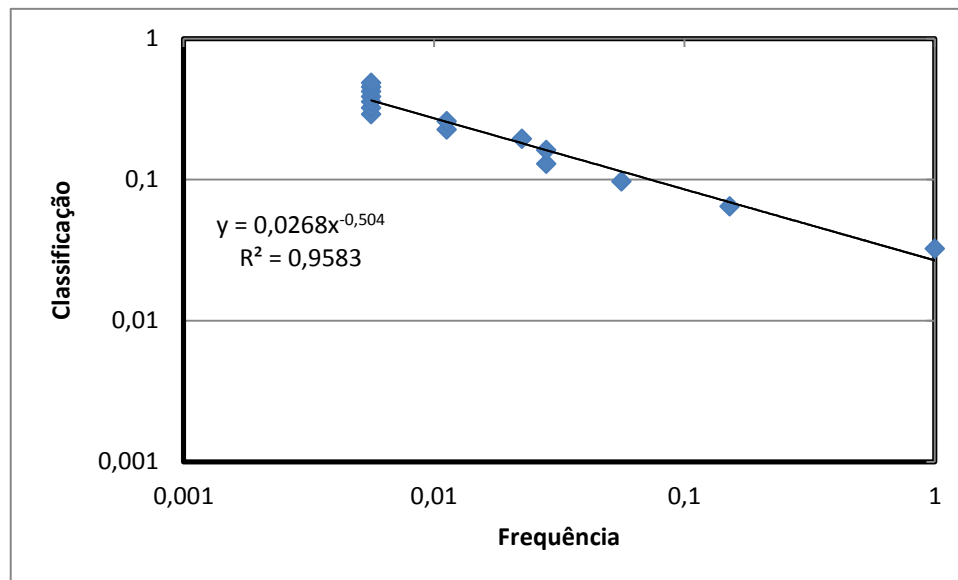


Figura 13 Gráfico classificação/frequência após normalização: 1972-1992

O período em questão foi marcado por uma menor intensidade de guerras, onde o número de vítimas se situou no intervalo [0,422]. De salientar que, das 255 guerras ocorridas neste período, apenas 22 tiveram um número de vítimas superiores a 10. Os dados são bem aproximados por uma recta, no gráfico log-log, pelo que se pode concluir que seguem uma distribuição LP (Figura 13).

3.4.2 Período de 1992-2002

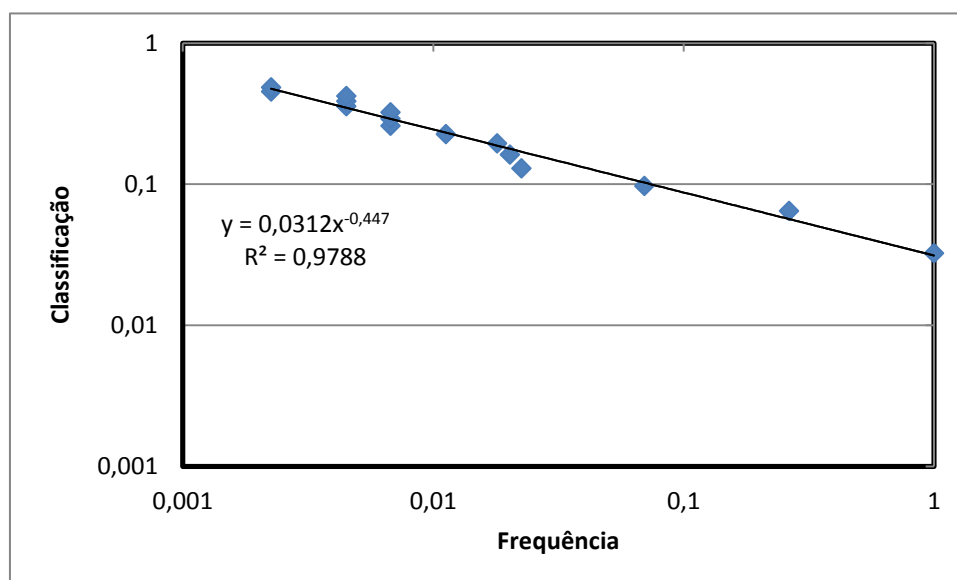


Figura 14 Gráfico classificação/frequência após normalização: 1992-2002

No período de 1992-2002, o número de vítimas situa-se no intervalo [0,1382]. Das 659 guerras analisadas apenas 31 causaram um número de vítimas superior a 10. Na Figura 14 observamos uma boa aproximação dos dados, no gráfico log-log, por uma recta, sendo o valor do coeficiente de determinação próximo de 1.

3.4.3 Período de 2002-2012

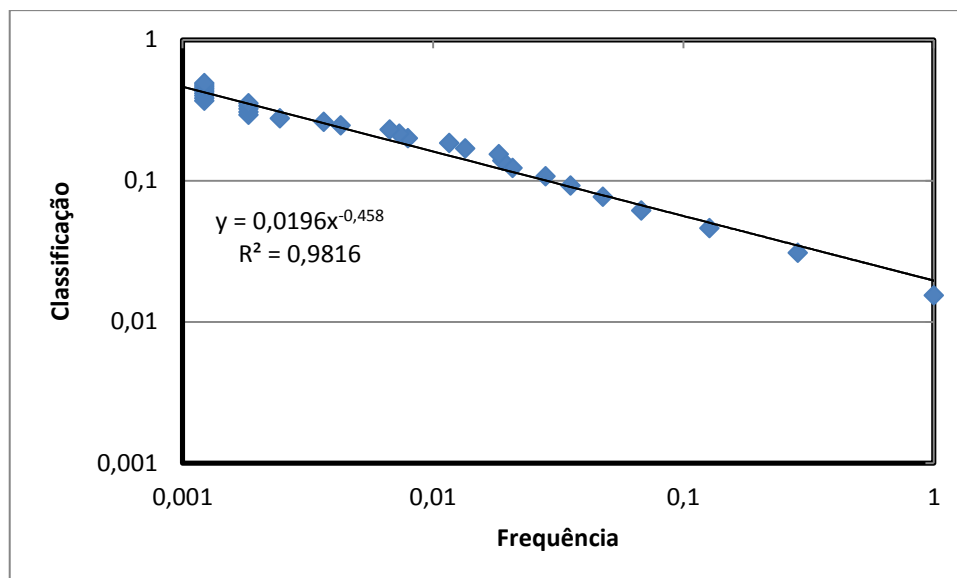


Figura 15 Gráfico classificação/frequência antes da normalização: 2002-2012

O último período em análise, de 2002 a 2012, incorpora a maioria dos dados (2860 dados nesta série temporal). Os dados são melhor aproximados pela recta de regressão, num gráfico log-log, sendo o coeficiente de determinação mais próximo de 1, comparado com períodos anteriores estudados acima. Neste caso os dados seguem mais fielmente uma distribuição LP (ver Figura 15).

Na Figura 16 podemos observar a relação entre os parâmetros das LPs para cada período estudado.

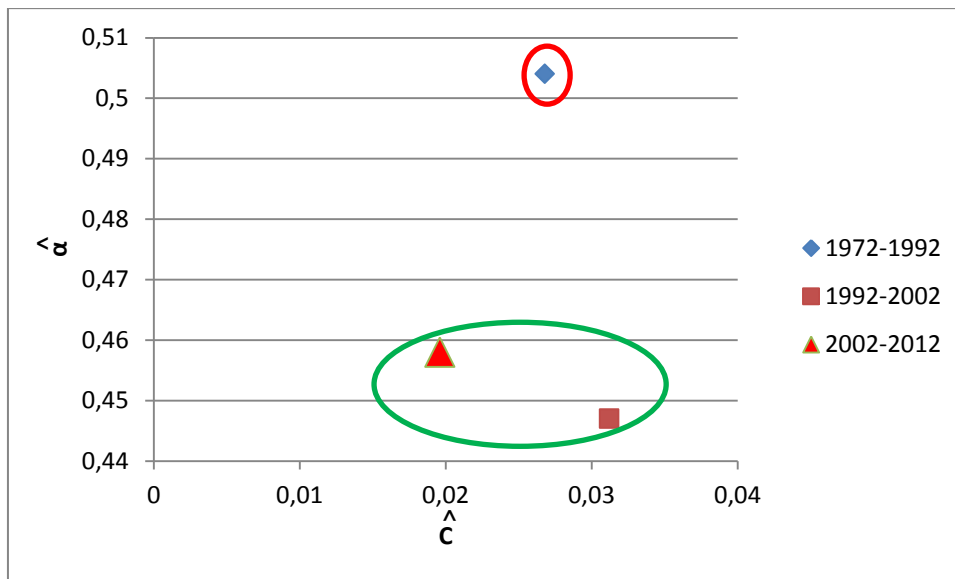


Figura 16 Representação dos parâmetros \hat{C} e $\hat{\alpha}$ por período de tempo

Os valores calculados de \hat{C} satisfazem a relação $0.0196 \leq \hat{C} \leq 0.0312$, onde os valores mínimos e máximos correspondem aos períodos 2002-2012 e 1992-2002, respectivamente. No que respeita ao parâmetro $\hat{\alpha}$ os valores estimados encontram-se entre $0.447 \leq \hat{\alpha} \leq 0.504$, onde os valores mínimos e máximos correspondem aos períodos 1992-2002 e 1972-1992, respectivamente. Em conclusão, distinguimos 2 zonas no gráfico, uma correspondente aos períodos de 1992-2002 e 2002-2012 (delimitado a verde) e outra ao período de 1972-1992 (a vermelho). Os períodos na zona a verde diferenciam-se de um período mais severo, em termos de número de vítimas, a vermelho (1972-1992). Observa-se uma diminuição da severidade das guerras desde 1992, com uma redução no valor de $\hat{\alpha}$.

3.5. VARIAÇÃO TEMPORAL DOS PARÂMETROS DAS LP

Nesta secção analisamos a variação dos parâmetros \hat{C} e $\hat{\alpha}$ ao longo do tempo, de modo a poder identificar padrões de comportamento e possivelmente indicar tendências que as

guerras possam seguir futuramente. A análise desta variação foi feita apenas no período de 1990 a 2012, devido à escassez de dados de anos anteriores.

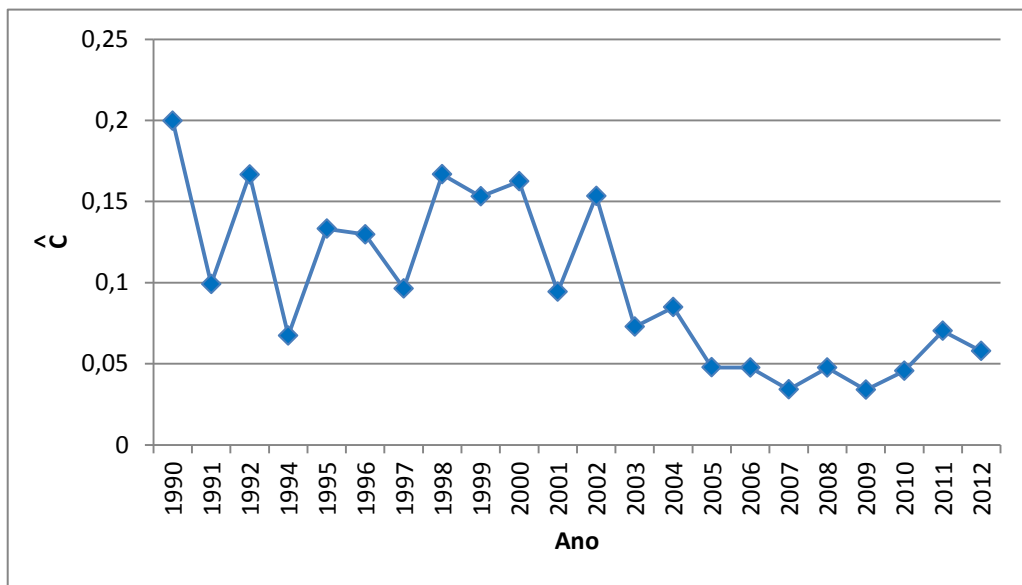


Figura 17 Evolução temporal do parâmetro \hat{C}

Observa-se uma tendência geral decrescente do valor de \hat{C} , existindo, todavia, algumas flutuações ao longo do tempo (Figura 17).

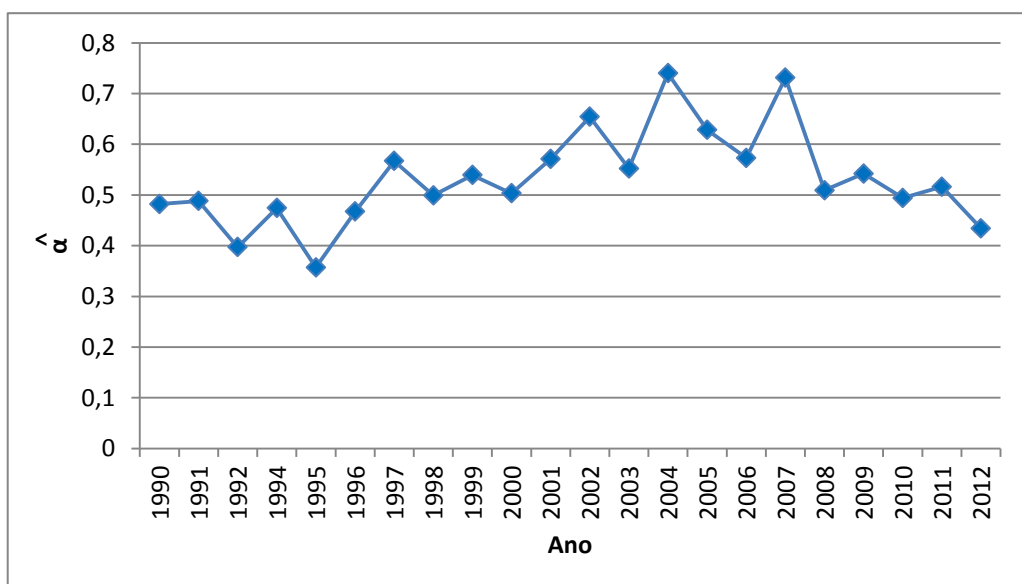


Figura 18 Evolução temporal do parâmetro $\hat{\alpha}$

No que diz respeito à evolução temporal do parâmetro $\hat{\alpha}$, observa-se uma tendência de crescimento, com picos em 2004 e 2007, sofrendo um decréscimo, a partir de 2007, para valores muito próximos dos de 1990 (ver Figura 18).

As conclusões gerais que se podem retirar da análise da evolução temporal dos dois parâmetros são: valores baixos de parâmetro \hat{C} e elevados de $\hat{\alpha}$ estão associados a uma maior severidade de guerras (período de 2004 a 2007). A partir de 2007, observa-se um decréscimo do valor do parâmetro $\hat{\alpha}$ e uma ligeira subida do valor do parâmetro \hat{C} . Isto pode representar uma tendência para uma menor severidade das guerras a partir de 2007. No entanto, estas conclusões não podem ser extrapoladas para o futuro, dado que há outros factores, como factores políticos, económicos, sociais, extraordinariamente importantes, que podem alterar dramaticamente este cenário.

4. CONCLUSÕES

Neste trabalho estudam-se distribuições estatísticas LP e a sua aplicação a guerras e actos de terrorismo. Aplicam-se as LP na análise de dados reais sobre vítimas de guerras e actos de terrorismo a nível mundial. Os dados são estudados por região, por período e por tipologia de guerra. Faz-se ainda uma análise da evolução temporal dos valores dos parâmetros das LP.

Os resultados obtidos permitem-nos concluir que todos os dados estudados são bem aproximados por uma distribuição LP. Isto é, o gráfico da frequência vs classificação, em dupla escala logarítmica, é bem aproximado por uma recta com declive negativo.

É possível identificar as assimetrias a nível regional do número de vítimas. A região da Ásia está associada a mais severidade, com maior frequência de vítimas. Inversamente, estão a Europa e a Austrália. Quanto à tipologia das guerras, salienta-se que existem mais guerras do tipo guerrilhas do que de carácter religioso, sendo estas mais severas do que as anteriores e estando-lhes associados valores mais baixos de \hat{C} . Relativamente ao período histórico, observa-se uma diminuição da severidade das guerras desde 1992, com uma redução no valor de $\hat{\alpha}$. Por fim é apresentada a evolução temporal dos parâmetros \hat{C} e $\hat{\alpha}$. O parâmetro \hat{C} apresenta uma diminuição global ao longo do tempo, havendo um ligeiro

crescimento no ano de 2011. O parâmetro $\hat{\alpha}$ sofreu globalmente um aumento de 1995 a 2007 e tem vindo a decrescer.

Os resultados obtidos precisam ser fundamentados teoricamente num trabalho futuro isto é, usar-se-ão outras distribuições, como a exponencial ou a lognormal, com comportamentos semelhantes à LP, com o objectivo de rejeitá-las comparativamente à LP e assim reforçar as conclusões. Dever-se-ão ainda aferir da influência do comportamento do ser humano, das motivações políticas, religiosas e étnicas, do nível de vida das populações, na distribuição das vítimas de guerras e actos de terrorismo. Deverá ser também alargado o período histórico analisado.

Referências Documentais

- [1] Auerbach F., “Das Gesetz der Bevölkerungskonzentration”, *Petermanns Geographische Mitteilungen*, vol. 59, pp. 74–76 (1913).
- [2] Bauke H., “Parameter estimation for power-law distributions by maximum likelihood methods”, *The European Physical Journal B*, vol. 58, no. 2, pp. 167-173 (2007).
- [3] Bogen K. T. and Jones E. D., “Risks of mortality and morbidity from worldwide terrorism: 1968–2004” *Risk Analysis*, vol. 26, no. 1, pp. 45–59 (2006).
- [4] Bohorquez J. C., Gourley S., Dixon A. R., Spagat M., Johnson N. F., “Common ecology quantifies human insurgency”, *Nature*, vol. 462, no. 7275, pp. 911–4 (2009).
- [5] California Department of Forestry and Fire Protection, URL: <http://www.fire.ca.gov/>.
- [6] Cederman L. E., “Modeling the size of wars: from billiard balls to sandpiles” *American Political Science Review*, vol. 97, pp. 135–50 (2003).
- [7] City Population, population statistics for countries, administrative areas, cities and agglomerations, URL: <http://www.citypopulation.de>.
- [8] Clauset A., Shalizi C., Newman M. E. J., “Power-Law Distributions in Empirical Data”, *SIAM Review*, no. 51, pp. 661-703 (2009).
- [9] Clauset A., Young M. and Gleditsch K. S., “On the frequency of severe terrorist events”, *Journal of Conflict Resolution*, vol. 51, no.1, pp. 58–87 (2007).
- [10] Ehlers R., “Inferência Estatística (Notas de Aula)”, Departamento de Matemática Aplicada e Estatística USP (2011).
- [11] Geller D. S. and Singer J. D, *Nations at War: A Scientific Study of International Conflict*, Cambridge Studies in International Relations, Cambridge University Press, Cambridge, U.K. (1998).
- [12] Hilborn R. and Mangel M., “The Ecological Detective: Confronting Models with Data (Monographs in Population Biology 28)”, Princeton University Press, Princeton (1997).
- [13] Jiménez Y. A. and Gallego J., “Desigualdad y leyes de potencia”, *Cuadernos de Economía*, vol. XXIX, no. 53, pp. 57-95 (2010).
- [14] Johnson N.F., Spagat M., Restrepo J. A., Becerra O., Bohorquez J. C., Suarez N., Restrepo E. M., Zarama R., “Universal patterns underlying ongoing wars and terrorism”, <http://arxiv.org/abs/physics/0605035> (2006).
- [15] KernScores, library of virtual musical scores, <http://kern.humdrum.org>.

- [16] Monteiro J. F. G., “Estimadores de máxima verosimilhança para a combinação captura – recaptura e trajectos lineares e as suas propriedades”, Universidade de Évora (2001).
- [17] Multimedia Library, electronic dictionary of musical themes, URL: <http://www.multimedialibrary.com/barlow>.
- [18] Newman M. E. J., “Power Laws, Pareto Distributions and Zipf’s Law”, *Contemporary Physics*, vol. 46, pp 323-351 (2005).
- [19] Pareto V., *Cours d’Economie Politique*, Droz, Geneva, Switzerland, 1896.
- [20] Pinto C. M. A., Lopes A. M., Tenreiro Machado J. A., “A review of power laws in real live phenomena”, *Communications in Nonlinear Science and Numerical Simulation*, vol. 17, pp. 3558-78 (2012).
- [21] Pinto C. M. A., Lopes A. M., Tenreiro Machado J. A., “Pareto and Zipf laws for city size distribution”, *Proceedings of the FSS 2011 - Symposium on Fractional Signals and Systems*, (2011). ISBN: 978--989--8331--13--7.
- [22] Portal Action, Método da Máxima Verosimilhança, URL: <http://www.portalaction.com.br/1310-o-m%C3%A9todo-de-m%C3%A1xima-verossimilhan%C3%A7a>.
- [23] Rafailidis D. and Manolopoulos Y., “The Power of Music: Searching for Power-Laws in Symbolic Musical Data”, *Proceedings of 12th Panhellenic Conference on Informatics (PCI’08)*, pp. 93-104 (2008).
- [24] Richardson L. F., “Variation of the frequency of fatal quarrels with magnitude”, *Journal of the American Statistical Association*, vol. 43, pp. 523–46 (1948).
- [25] Roberts D. C. and Turcotte D. L., “Fractality and selforganized criticality of wars”, *Fractals*, vol. 6, pp. 351–7 (1998).
- [26] Sasse F. D., Departamento de Matemática CCT- UDESC, URL: http://deeke.org/point_estimation.pdf.
- [27] Small M. and Singer J. D., “Resort to Arms: International and Civil Wars, 1816-1980”, Sage Publications, Beverley Hills (1982).
- [28] Tenreiro Machado J. A., Pinto C. M. A., Lopes A. M., “Power Law and Entropy Analysis of Catastrophic Phenomena”, *Mathematical Problems in Engineering*, vol. 2013, no. 28, pp. 1-10 (2013).
- [29] Wolfram Mathworld, Least Squares Fitting – Power Law, URL: <http://mathworld.wolfram.com/LeastSquaresFittingPowerLaw.html>.
- [30] Zipf G., “Human Behavior and the Principle of Least Effort”, Addison-Wesley, Cambridge, Mass, USA, 1949.